

Can the sample variance estimator be improved by using a covariate?

Hans-Peter Piepho

Senior Statistician, Institut für Nutzpflanzenkunde, Universität-Gesamthochschule Kassel,
Steinstrasse 19, 37213 Germany

Charles E. McCulloch

Professor of Biological Statistics, Department of Biometrics, College of Agriculture and Life
Science, Cornell University, 434 Warren Hall, Ithaca, NY 14853, USA

Variance components are quantities of central interest in many applications, for example in cultivar yield stability analysis and in the analysis of measurement errors. In some applications, the feasible sample size is rather limited leading to estimates of variance components which are subject to considerable sampling variation. For example, new crop cultivars are tested for only a few years before release to the market, so the sample size for the year variance component is small. Similarly, testing a new measurement instrument for some chemical compound may be costly, allowing only a limited number of replications. This paper investigates the potential for improving the usual sample variance estimator by exploiting covariate information. In a cultivar trial, yield data may be available for only a few environments, while meteorological data or data on a standard cultivar has been recorded for a very large number of environments. Likewise, in the analysis of measurement errors, there may be long-term data on a standard measurement procedure that can be used as a covariate to improve the variance estimate for a new instrument. It is shown in this paper that the gain in accuracy by a covariate can be considerable, provided there is sufficient correlation between the covariate and the variable of interest.

Key Words: Bivariate normality, covariates, mean square error (MSE), variance component, restricted maximum likelihood (REML), unbiasedness, yield stability.

1. Introduction

Consider a variable of interest y_i ($i = 1, \dots, K$) and an associated covariate t_i . Assume that $\mathbf{z}_i = (y_i, t_i)'$ has a bivariate normal distribution (BVN), i.e. $\mathbf{z}_i \sim \text{BVN}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$, where $\boldsymbol{\mu}_z = (\mu_y, \mu_t)'$ and

$\boldsymbol{\Sigma}_z = \begin{pmatrix} \sigma_y^2 & \sigma_{yt} \\ \sigma_{ty} & \sigma_t^2 \end{pmatrix}$. The model for y_k may be written as a simple linear regression

$$y_i = \theta + \beta t_i + e_i \quad (1)$$

where $\theta = \mu_y - \beta \mu_t$, $\beta = \sigma_{yt} \sigma_t^{-2}$ and e_i is a random normal deviate, which is independent of t_i and which has zero expectation and variance $\sigma_e^2 = \sigma_y^2 - (\sigma_{yt})^2 \sigma_t^{-2}$. The variance of y_i can be rewritten as

$$\sigma_y^2 = \beta^2 \sigma_t^2 + \sigma_e^2 \quad (2)$$

We are interested here in estimation of σ_y^2 . The question to be answered is whether the usual estimator

$$\hat{\sigma}_y^2 = a \sum_{i=1}^K (y_k - \bar{y})^2 \quad (3)$$

with $a = a_{unb} = (K - 1)^{-1}$ can be improved upon by exploiting the correlation of y_i and t_i . One might conjecture from inspection of (2) that an improvement is possible when good estimates of σ_t^2 and β are available.

This study was triggered by work on across-year stability of yield in cultivar trials. In this context, y_i is the yield of a cultivar in the i -th year at a given location and t_i is a year-specific covariate, for example the average temperature. The variance component σ_y^2 is a measure of the stability of yield across years (Piepho, 1998). A common feature of cultivar trials is that the number of years is limited, typically ranging from three to ten. In this situation, estimates of σ_y^2 based on (3) are extremely unreliable. The conjecture motivating this study was that estimation of σ_y^2 can be improved by using long-term data t_i on a meteorological covariate or

on a standard cultivar. The long-term data can provide a highly accurate estimate of σ_t^2 .

Another potential field of application is the analysis of measurement errors (Jaech, 1985; Fuller, 1987). The accuracy (variance) of a standard measurement procedure may have been measured so often that it can be considered as a known quantity. If measurement errors of a new procedure and the standard are reasonably correlated, a few paired observations on both procedures may be sufficient to obtain a good estimate for the new procedure by exploiting long-term information on the standard. Similarly, if testing of a new measurement device is very costly relative to a standard, making additional observations on the standard can provide a cost-effective means of increasing the accuracy of the variance estimate for the new device.

2. Variance estimators

It is assumed that for the quantity of interest data are available for $i = 1, \dots, K$, while covariate information is available for $i = 1, \dots, M$ where $M \geq K$:

$$y = (y_1, y_2, \dots, y_K)' \quad (4)$$

$$t = (t_1, t_2, \dots, t_K, t_{K+1}, \dots, t_M)' \quad (5)$$

For example, in addition to cultivar yield data y_i there may be long-term data t_i on a standard cultivar or on average yearly temperatures, so that $M \gg K$. If the covariate information is ignored, we use the sample variance in eq. (3). If the covariate information is available in the years $i = 1, \dots, M$ and if $K > 2$, we consider estimators of the form

$$\tilde{\sigma}_y^2 = b \hat{\beta}^2 \hat{\sigma}_t^2 + c \hat{\sigma}_e^2 \quad (6)$$

where

$$\hat{\beta} = SS_1^{-1} CP,$$

$$SS_1 = \sum_{i=1}^K (t_i - \bar{t}^{(1)})^2,$$

$$CP = \sum_{i=1}^K (y_i - \bar{y}_\cdot)(t_i - \bar{t}_\cdot^{(1)}) ,$$

$$\bar{y}_\cdot = K^{-1} \sum_{i=1}^K y_i ,$$

$$\bar{t}_\cdot^{(1)} = K^{-1} \sum_{i=1}^K t_i ,$$

$$\hat{\sigma}_t^2 = (M - 1)^{-1} SS_M ,$$

$$SS_M = \sum_{i=1}^M (t_i - \bar{t}_\cdot)^2 ,$$

$$\bar{t}_\cdot = M^{-1} \sum_{i=1}^M t_i ,$$

$$\hat{\sigma}_e^2 = (K - 2)^{-1} \sum_{i=1}^K (y_i - \hat{\theta} - \hat{\beta} t_i)^2 ,$$

$$\hat{\theta} = \bar{y}_\cdot - \hat{\beta} \bar{t}_\cdot^{(1)} ,$$

and b and c are a suitably chosen constants, possibly depending only on K and M . Use of this type of estimator is motivated by the fact that the restricted maximum likelihood (REML) estimator of σ_y^2 based on the data (y, t) is of the form (6) with $b = 1$ and c equal to

$$c = c_{REML} = (K - 2)(K - 1)^{-1} ; \quad K > 2 \quad (7)$$

(see Appendix D). This estimator is biased. Unbiasedness is attained by choosing (see Appendix B) $b = 1$ and

$$c = c_{unb} = 1 - (K - 3)^{-1}(M - 1)^{-1}(M - 3) ; \quad K > 3 \quad (8)$$

Note that the REML estimator ($b = 1$ and c_{REML}), the unbiased estimator ($b = 1$ and c_{unb}) and the simple estimator (3) with $a = a_{unb}$ coincide when $K = M$ ($K > 3$).

The choice $b = 1$ follows from the requirement of unbiasedness. If we do not require unbiasedness, minimization of the mean squared error (MSE) is an appropriate optimality criterion. Unfortunately, the minimum MSE solution for b and c depends on the parameters (Appendix C), so minimum MSE estimation is not generally feasible. Note that dependence of minimum MSE estimation is a general problem in many applications (Stuart and Ord, 1991, §

17.30), which is the main reason for its limited use. Due to the unfeasibility of an overall minimum MSE estimator, we will focus on the unbiased estimators.

3. The mean squared error of estimators

Since some of the proposed estimators are biased, i.e. the REML estimator and the minimum MSE estimators discussed in Appendix C, efficiency will be compared by the MSE. Note that for unbiased estimators, the MSE is equivalent to the variance. The simple estimator (3) has MSE (Cox and Hinkley, 1974, § 8.5)

$$\text{MSE}(\hat{\sigma}_y^2) = \{2a^2(K-1) + [a(K-1) - 1]^2\} \sigma_y^4 \quad (9)$$

It is shown in Appendix C that $\tilde{\sigma}_y^2$ in (6) has MSE

$$\begin{aligned} \text{MSE}(\tilde{\sigma}_y^2) = & [2b^2(M-1)^{-1} + (b-1)^2] \beta^4 \sigma_t^4 \\ & + [6b^2(M-1)^{-2}F - 2(M-1)^{-1}D + 2(b-1)(c-1)] \beta^2 \sigma_t^2 \sigma_e^2 \\ & + [3b^2(M-1)^{-2}E + 2c^2(K-2)^{-1} + 2b(c-1)(M-1)^{-1}D + (c-1)^2] \sigma_e^4 \end{aligned} \quad (10)$$

where

$$\begin{aligned} D &= (M-3)(K-3)^{-1} \quad ; \quad K > 3 \\ E &= 1 + 2(M-K)(K-3)^{-1} + (M-K+2)(M-K)(K-3)^{-1}(K-5)^{-1} \quad ; \quad K > 5 \\ F &= (K-1) + 2(M-K) + (K-3)^{-1}(M-K+2)(M-K) \quad ; \quad K > 3 \end{aligned} \quad (11)$$

This equation can be used to compare different estimators. The quantities D , E and F are expectations of ratios of powers of SS_M and SS_I . These expectations exist only for $K > 3$, $K > 5$ and $K > 3$, respectively (Appendix A). Thus, the MSE can be studied only for $K > 5$. Note that similar results hold for the F-distribution (Johnson and Kotz, 1970b, Chapter 26). For smaller K , other criteria would have to be used, for example the interquartile range, but this is not pursued here for brevity.

For the case of known variance component σ_t^2 , simply replace $\hat{\sigma}_t^2$ by σ_t^2 in (6) and let $M \rightarrow \infty$ in (8), (9), and (10). We find for $M \rightarrow \infty$

$$c_{unb} = (K - 4)(K - 3)^{-1} ; \quad K > 3 \quad (12)$$

and for $b = 1$ and $K > 5$

$$\begin{aligned} \text{MSE}(\tilde{\sigma}_y^2) &= 4(K - 3)^{-1} \beta^2 \sigma_t^2 \sigma_e^2 \\ &+ [3(K - 3)^{-1}(K - 5)^{-1} + 2c^2(K - 2)^{-1} + 2(c - 1)(K - 3)^{-1} + (c - 1)^2] \sigma_e^4 \end{aligned} \quad (13)$$

4. Comparison of estimators

When unbiasedness is insisted upon and minimum variance estimation is the objective, the estimators $\hat{\sigma}_y^2$ in (3) with $a = a_{unb}$ and $\tilde{\sigma}_y^2$ in (6) with $b = 1$ and $c = c_{unb}$ are available. These two estimators will be studied here in some detail. As an index of performance we use the quantity

$$P = \text{MSE}(\tilde{\sigma}_y^2) / \text{MSE}(\hat{\sigma}_y^2). \quad (15)$$

This index depends on K and M as well as on the ratio

$$q = \sigma_e^2 / (\beta^2 \sigma_t^2 + \sigma_e^2), \quad (16)$$

the fraction of variation in y that is not 'explained' by covariation with t . It is assumed throughout that $M > K > 5$.

When $q = 0$, then y and t are perfectly correlated, so that β can be measured without error, and observations t_k ($k = K + 1$ to M) are equivalent to additional observations y_k . We find

$$\text{MSE}(\tilde{\sigma}_y^2) = 2(M - 1)^{-1} \sigma_y^2, \text{ so } P = (K - 1)(M - 1)^{-1} < 1 \text{ for } M > K \text{ and use of covariate}$$

information is always worthwhile. A dramatic gain is possible, when $M \gg K$. By contrast, for $q = 1$ it can be shown by tedious but straightforward calculations (supported by using

Mathematica) that $P < 1$ always. This is expected, since in the absence of correlation, incorporation of a covariate into an estimator cannot provide any gain. Now P is quadratic in q , which has constant second derivative. Thus, there is always exactly one value of q , q_P say ($0 < q_P < 1$), so that $P > 1$ for $q < q_P$ and $P < 1$ for $q > q_P$. Hence, for any K and M , for which $M > K > 5$, there is a point q_P below which it is always worthwhile to exploit covariate information. The gain in accuracy increases with decreasing value of q .

To gain further insight, consider the following first order expansion (ignoring a covariance term)

$$\text{var}(\hat{\beta}^2 \hat{\sigma}_t^2) \sigma_y^{-4} \approx T_1 + T_2, \quad (17)$$

where $T_1 = [E(\hat{\sigma}_t^2)]^2 \text{var}(\hat{\beta}^2) \sigma_y^{-4}$ and $T_2 = [E(\hat{\beta}^2)]^2 \text{var}(\hat{\sigma}_t^2) \sigma_y^{-4}$. The term T_2 in (17) can be reduced by increasing M , while the first term cannot. It can be shown algebraically (Appendix E) that $\text{var}(\hat{\beta}^2 \hat{\sigma}_t^2)$ will always be dominated by T_1 (i.e. $T_1 > T_2$) and hence by $\text{var}(\hat{\beta}^2)$, when $q > 1/3$. When $q < 1/3$, (17) will usually also be dominated by T_1 unless q is rather small and/or M is close to K . This shows that variability in the estimate of β^2 is the Achilles heel of our procedure. Unless the correlation between y and t is perfect, a certain number of paired observations (K) may be necessary before enough accuracy is gained so that exploitation of the covariate becomes worthwhile. If the correlation is too weak (q too large), it may not be possible to achieve any gain at all, even when M is large. Fig. 1 shows this for $M = 100$.

Fig. 2 show plots of P vs. q , for K equal to 6, 10 and 20 and $M = 30$. There is a notable shift of the curve for increasing K . For $K = 6$ the crossing point q_P is at about 0.3, and above that point P becomes very unfavorable. For $K = 20$ the crossing point shifts to $q_P \approx 0.75$ and above that point P remains very close to 1. This pattern remains fairly stable across different values of M (see Fig. 3). Thus, the value of K is critical to the performance of the estimator (6). The main problem with small K appears to be poor estimation of β . It can be concluded that for larger K one does not lose much and will most likely win by generally exploiting covariate information, while for small K , knowledge of q is crucial for an appropriate choice of estimator and the covariate should not be used for large q .

Very similar results are found for other estimators of the form (6). As an example, Fig. 4

shows a plot of P vs. q for $M = 30$, $b = 1$ and $c = c_{reml}$ with different values of K . The graphs are very close to those in Fig. 1 for the unbiased estimator. For the constrained minimum MSE estimators discussed in Appendix C, we also found similar results, but these are not shown for brevity and because we favor the unbiased estimator due to unfeasibility of unconstrained minimum MSE estimation.

5. Numerical example

We use yield data from international wheat trials conducted by CIMMYT (Centro International del Mejoramiento del Maiz y Trigo, Mexico) from 1975 to 1986 to exemplify the method. We use an unbalanced subset of a larger data set that has been analyzed by Piepho and van Eeuwijk (1999). Data on two genotypes, one standard (t_k) and one new cultivar (y_k), in one location are displayed in Table 1.

We use the first cultivar as a covariate (t_k) for estimating the variance of the second cultivar (y_k). Using the unbiased estimators in (3) and (6) with $M = 12$ and $K = 6$, we find $\hat{\sigma}_y^2 = 50.75$ for the covariate-free estimator (3) and $\tilde{\sigma}_y^2 = 62.22$ for the estimator based on the covariate (6) ($\hat{\beta} = 1.049$, $\hat{\sigma}_t^2 = 43.84$, $\hat{\sigma}_e^2 = 19.19$). The estimated fraction of variation not explained by the covariance between y_k and t_k is $\hat{q} = \hat{\sigma}_e^2 / \tilde{\sigma}_y^2 = 0.308$, which is rather low. Plugging these estimates into eq. (8) and (9) we obtain $MSE(\hat{\sigma}_y^2) = 0.400$ and $MSE(\tilde{\sigma}_y^2) = 0.387$. Thus, use of the covariate seems to be useful, although we have to keep in mind that variance component estimates were used to estimate the MSE .

6. Concluding remarks

It is known from linear regression that predictive accuracy of a full regression model may be inferior to a reduced model, when the regressor variables do not explain a large enough share of the total variation in the response (Hocking, 1976; Piepho, Denis, and van Eeuwijk, 1998). This same behavior was observed in the present paper for estimation of a variance component in the presence of covariate information.

In cultivar trial data, it is often found, that the correlation between cultivar yield and covariates is weak, so that q is large. Also, a pressing problem in practice is to improve variance estimation when K is small. Unfortunately, it is exactly in these circumstances, that use of covariate information offers little if any improvement over the usual variance estimator (3), unless the correlation is high. If, however, K is not too small and the correlation is reasonably tight, use of a covariate can be very worthwhile. Usually, several cultivars are tested along-side each other. Instead of using a physical measurement on the environment as a covariate, one can use yield data of a standard variety or the average of several standards, for which long-term data are available. Correlation among cultivars is usually large, so the potential gain in accuracy can be substantial.

In this paper we have focused on yield stability across years as an example. Many studies assess stability in terms of variance across locations. Since in multilocation trials, the number of environments K is usually quite large (>20), use of covariate information for estimating across-location variance is almost always worthwhile.

Confidence limits for σ_y^2 based on $\tilde{\sigma}_y^2$ in (6) with $b = 1$ and $c = c_{unb}$ can be obtained using a Satterthwaite-approximation. Specifically, it is conjectured that $\tilde{\sigma}_y^2$ approximately follows a χ^2 -distribution with degrees of freedom (df) equal to

$$df \approx \frac{2[E(\tilde{\sigma}_y^2)]^2}{\text{var}(\tilde{\sigma}_y^2)} \quad (18)$$

To evaluate (18), estimators of have to be inserted for the variance parameters. The behavior of this approximation needs to be evaluated by simulation in future work.

Acknowledgements

We are grateful to the CIMMYT Regional Office for Latin America and INIA-Uruguay for allowing us to use the data. Special thanks go to Sergio Ceretta (INIA-Uruguay) for compiling

the data we have used. Part of the research for this paper was conducted while the first author was visiting the Department of Biometrics, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, USA. Support of the Heisenberg Programm of the Deutsche Forschungsgemeinschaft (DFG) and NSF DMS 9625476 is thankfully acknowledged.

References

- Cox, D. R., and Hinkley, D. V. (1974), *Theoretical statistics*, London: Chapman & Hall, London.
- Fuller, W. A. (1987), *Measurement error models*. New York: Wiley.
- Hocking, R. R. (1976), "The analysis and selection of variables in linear regression," *Biometrics*, **32**, 1-10.
- Jaech, J. L. (1985), *Statistical analysis of measurement errors*, New York: Wiley.
- Johnson, N. L., and Kotz, S. (1970a), *Continuous univariate distributions 1*, New York: Wiley.
- Johnson, N. L., and Kotz, S. (1970b), *Continuous univariate distributions 2*. New York: Wiley.
- Piepho, H. P. (1998), "Methods for comparing the yield stability of cropping systems - A review," *Journal of Agronomy and Crop Science*, **180**, 193-213.
- Piepho, H. P., Denis, J. B., and van Eeuwijk, F. A. (1998), "Predicting cultivar differences using covariates," *Journal of Agricultural, Biological and Environmental Statistics*, **3**, 151-162.
- Searle, S. R. (1982), *Matrix algebra useful for statistics*, New York: Wiley.
- Searle, S. R. (1987), *Linear models for unbalanced data*, New York: Wiley.
- Searle, S. R., Casella, G., and McCulloch, C. (1992), *Variance components*, New York: Wiley.
- Stuart, A., and Ord, K. (1991), *Kendall's Advanced theory of statistics*, Volume 2, London Edward Arnold.

Appendix A (Appendices will be shortened later; see marked passages. Extended version is intended for reviewers)

In Appendices B and C we will need expectations of ratios involving SS_1 and SS_M . To derive these expectations, first note that (Searle, 1987, p. 31)

$$SS_M = SS_1 + SS_2 + SS_3, \quad \text{where}$$

$$SS_2 = \sum_{i=K+1}^M (t_i - \bar{t}^{(2)})^2 \quad \text{with} \quad \bar{t}^{(2)} = (M-K)^{-1} \sum_{i=K+1}^M t_i \quad \text{and}$$

$$SS_3 = K(\bar{t}^{(1)} - \bar{t})^2 + (M-K)(\bar{t}^{(2)} - \bar{t})^2$$

The quantities SS_1/σ^2 , SS_2/σ^2 , and SS_3/σ^2 are stochastically independent χ^2 -variables with, respectively, $(K-1)$, $(M-K-1)$, and one degrees of freedom. From properties of the χ^2 -distribution (Johnson and Kotz, 1970a, Chapter 17)

$$E(SS_M) = (M-1)\sigma_t^2$$

$$E(SS_M^2) = (M+1)(M-1)\sigma_t^4$$

$$E(SS_K) = (K-1)\sigma_t^2$$

$$E(SS_1^{-1}) = (K-3)^{-1}\sigma_t^{-2} \quad (K > 3; \text{expectation infinity otherwise})$$

$$E(SS_1^{-2}) = (K-3)^{-1}(K-5)^{-1}\sigma_t^{-4} \quad (K > 5; \text{expectation infinity otherwise})$$

$$E(SS_2 + SS_3) = (M-K)\sigma_t^2$$

$$E[(SS_2 + SS_3)^2] = (M-K+2)(M-K)\sigma_t^4$$

We rewrite $SS_M SS_1^{-1} = 1 + R$ where $R = (SS_2 + SS_3)SS_1^{-1}$. By the independence of SS_1 and $(SS_2 + SS_3)$ we have:

$$E(R) = E(SS_2 + SS_3)E(SS_1^{-1}) = (M-K)(K-3)^{-1} \quad \text{for } K > 3$$

$$E(R^2) = E[(SS_2 + SS_3)^2]E(SS_1^{-2}) = (M-K+2)(M-K)(K-3)^{-1}(K-5)^{-1} \quad \text{for } K > 5 \text{ and thus}$$

$$E(SS_M SS_1^{-1}) = 1 + E(R) = (M-3)(K-3)^{-1} = D \quad ; K > 3$$

$$E(SS_M^2 SS_1^{-2}) = 1 + 2E(R) + E(R^2) = 1 + 2(M-K)(K-3)^{-1} + (M-K+2)(M-K)(K-3)^{-1}(K-5)^{-1}$$

$$= E \quad ; K > 5$$

$$\begin{aligned} E(SS_M^2 SS_1^{-1}) &= \frac{E(SS_1) + 2E(SS_2 + SS_3) + E(SS_1^{-1})E[(SS_2 + SS_3)^2]}{[(K-1) + 2(M-K) + (K-3)^{-1}(M-K+2)(M-K)]\sigma_e^2} = F\sigma_e^2 \quad ; K > 3 \end{aligned}$$

Appendix B

We now derive an unbiased estimator of σ_y^2 for the case where covariate information is available in the years $k = 1, \dots, M$. Consider an estimator of the form

$$\tilde{\sigma}_y^2 = b\hat{\beta}^2\hat{\sigma}_t^2 + c\hat{\sigma}_e^2$$

Conditionally on $\mathbf{t} = (t_1, \dots, t_M)'$ we have

$$\begin{aligned} E(\hat{\beta}^2 | \mathbf{t}) &= \text{var}(\hat{\beta} | \mathbf{t}) + [E(\hat{\beta} | \mathbf{t})]^2 = SS_1^{-1}\sigma_e^2 + \beta^2, \text{ and thus} \\ E(\tilde{\sigma}_y^2 | \mathbf{t}) &= \frac{bE(\hat{\beta}^2 | \mathbf{t})\hat{\sigma}_t^2 + c\sigma_e^2}{b(M-1)^{-1}[SS_M SS_1^{-1}\sigma_e^2 + SS_M\beta^2]} + c\sigma_e^2 \\ &= b(M-1)^{-1}SS_M\beta^2 + [b(M-1)^{-1}SS_M SS_1^{-1} + c]\sigma_e^2 \end{aligned}$$

Using Appendix A we find

$$\begin{aligned} E(\tilde{\sigma}_y^2) &= E_t[E(\tilde{\sigma}_y^2 | \mathbf{t})] = \frac{b(M-1)^{-1}[E(SS_M SS_1^{-1})\sigma_e^2 + E(SS_M)\beta^2] + c\sigma_e^2}{b\beta^2\sigma_t^2 + [b(M-1)^{-1}D + c]\sigma_e^2} \\ &= b\beta^2\sigma_t^2 + [b(M-1)^{-1}D + c]\sigma_e^2 \end{aligned}$$

Unbiasedness results by setting $b = 1$ and $c = \frac{1 - (M-1)^{-1}D}{1 - (K-3)^{-1}(M-1)^{-1}(M-3)}$.

Appendix C

We here derive $\text{MSE}(\tilde{\sigma}_y^2)$. From results on conditional moments (Searle *et al.*, 1992, p.461)

$$\begin{aligned} \text{var}(\tilde{\sigma}_y^2) &= E_t[\text{var}(\tilde{\sigma}_y^2 | \mathbf{t})] + \text{var}_t[E(\tilde{\sigma}_y^2 | \mathbf{t})], \\ \text{var}_t[E(\tilde{\sigma}_y^2 | \mathbf{t})] &= E_t\{[E(\tilde{\sigma}_y^2 | \mathbf{t})]^2\} - \{E_t[E(\tilde{\sigma}_y^2 | \mathbf{t})]\}^2 = E_t\{[E(\tilde{\sigma}_y^2 | \mathbf{t})]^2\} - [E(\tilde{\sigma}_y^2)]^2, \text{ and} \end{aligned}$$

$$\begin{aligned}
\text{MSE}(\tilde{\sigma}_y^2) &= \text{var}(\tilde{\sigma}_y^2) + [\sigma_y^2 - E(\tilde{\sigma}_y^2)]^2 \\
&= E_t[\text{var}(\tilde{\sigma}_y^2 | \mathbf{t})] + E_t\{[E(\tilde{\sigma}_y^2 | \mathbf{t})]^2\} - [E(\tilde{\sigma}_y^2)]^2 + \sigma_y^4 - 2\sigma_y^2 E(\tilde{\sigma}_y^2) + [E(\tilde{\sigma}_y^2)]^2 \\
&= E_t[\text{var}(\tilde{\sigma}_y^2 | \mathbf{t})] + E_t\{[E(\tilde{\sigma}_y^2 | \mathbf{t})]^2\} + \sigma_y^4 - 2\sigma_y^2 E(\tilde{\sigma}_y^2)
\end{aligned}$$

Derivation of $E_t[\text{var}(\tilde{\sigma}_y^2 | \mathbf{t})]$: By the conditional independence of $\hat{\beta}$ and $\hat{\sigma}_e^2$, given \mathbf{t} , we have

$$\text{var}(\tilde{\sigma}_y^2 | \mathbf{t}) = \underline{b^2 \hat{\sigma}_t^4 \text{var}[(\hat{\beta})^2 | \mathbf{t}] + c^2 \text{var}(\hat{\sigma}_e^2 | \mathbf{t})} = b^2 (M-1)^{-2} SS_M^2 \text{var}[(\hat{\beta})^2 | \mathbf{t}] + c^2 \text{var}(\hat{\sigma}_e^2 | \mathbf{t})$$

Assuming normality, we have

$$\text{var}(\hat{\sigma}_e^2 | \mathbf{t}) = 2(K-2)^{-1} \sigma_e^4$$

From results on quadratic forms in normal random variables (Searle *et al.*, 1992, p. 467)

$$\begin{aligned}
\text{var}(\hat{\beta}^2 | \mathbf{t}) &= 2[\text{var}(\hat{\beta} | \mathbf{t})]^2 + 4[E(\hat{\beta} | \mathbf{t})]^2 \text{var}(\hat{\beta} | \mathbf{t}) = 2SS_1^{-2} \sigma_e^4 + 4\beta^2 SS_1^{-1} \sigma_e^2 \quad \text{and thus} \\
\text{var}(\tilde{\sigma}_y^2 | \mathbf{t}) &= \underline{b^2 (M-1)^{-2} SS_M^2 (2SS_1^{-2} \sigma_e^4 + 4\beta^2 SS_1^{-1} \sigma_e^2)} + 2c^2 (K-2)^{-1} \sigma_e^4
\end{aligned}$$

From Appendix A

$$\begin{aligned}
E_t[\text{var}(\tilde{\sigma}_y^2 | \mathbf{t})] &= \underline{b^2 (M-1)^{-2} [2E(SS_M^2 SS_1^{-2}) \sigma_e^4 + 4E(SS_M^2 SS_1^{-1}) \beta^2 \sigma_e^2]} + 2c^2 (K-2)^{-1} \sigma_e^4 \\
&= \underline{b^2 (M-1)^{-2} (2E \sigma_e^4 + 4F \beta^2 \sigma_t^2 \sigma_e^2)} + 2c^2 (K-2)^{-1} \sigma_e^4 \\
&= \underline{4b^2 (M-1)^{-2} F \beta^2 \sigma_t^2 \sigma_e^2} + 2[b^2 (M-1)^{-2} E + c^2 (K-2)^{-1}] \sigma_e^4
\end{aligned}$$

Derivation of $E_t\{[E(\tilde{\sigma}_y^2 | \mathbf{t})]^2\}$: From Appendix B

$$\begin{aligned}
[E(\tilde{\sigma}_y^2 | \mathbf{t})]^2 &= b^2 (M-1)^{-2} SS_M^2 \beta^4 + 2[b^2 (M-1)^{-2} SS_M^2 SS_1^{-1} + bc(M-1)^{-1} SS_M] \beta^2 \sigma_e^2 \\
&\quad + [b^2 (M-1)^{-2} SS_M^2 SS_1^{-2} + 2bc(M-1)^{-1} SS_M SS_1^{-1} + c^2] \sigma_e^4
\end{aligned}$$

Using results of Appendix A

$$\begin{aligned} E_t\{[E(\tilde{\sigma}_y^2 | t)]^2\} &= b^2(M+1)(M-1)^{-1} \beta^4 \sigma_t^4 + 2[b^2(M-1)^{-2}F + bc] \beta^2 \sigma_t^2 \sigma_e^2 \\ &\quad + [b^2(M-1)^{-2}E + 2bc(M-1)^{-1}D + c^2] \sigma_e^4 \end{aligned}$$

Putting results together:

$$\begin{aligned} -2 \sigma_y^2 E(\tilde{\sigma}_y^2) &= -2 \{\beta^2 \sigma_t^2 + \sigma_e^2\} \{b\beta^2 \sigma_t^2 + [b(M-1)^{-1}D + c]\sigma_e^2\} \\ &= -2 b\beta^4 \sigma_t^4 - 2[b(M-1)^{-1}D + c + b] \beta^2 \sigma_t^2 \sigma_e^2 - 2[b(M-1)^{-1}D + c] \sigma_e^4 \end{aligned}$$

so that

$$\begin{aligned} \text{MSE}(\tilde{\sigma}_y^2) &= E_t[\text{var}(\tilde{\sigma}_y^2 | t)] + E_t\{[E(\tilde{\sigma}_y^2 | t)]^2\} + \sigma_y^4 - 2 \sigma_y^2 E(\tilde{\sigma}_y^2) \\ &= [b^2(M+1)(M-1)^{-1} + 1 - 2b] \beta^4 \sigma_t^4 \\ &\quad + \{6b^2(M-1)^{-2}F + 2bc + 2 - 2[b(M-1)^{-1}D + c + b]\} \beta^2 \sigma_t^2 \sigma_e^2 \\ &\quad + \{2b^2(M-1)^{-2}E + 2c^2(K-2)^{-1} + b^2(M-1)^{-2}E + 2bc(M-1)^{-1}D + c^2 + 1 \\ &\quad \quad - 2[b(M-1)^{-1}D + c]\} \sigma_e^4 \\ &= [2b^2(M-1)^{-1} + (b-1)^2] \beta^4 \sigma_t^4 \\ &\quad + [6b^2(M-1)^{-2}F - 2b(M-1)^{-1}D + 2(b-1)(c-1)] \beta^2 \sigma_t^2 \sigma_e^2 \\ &\quad + [3b^2(M-1)^{-2}E + 2c^2(K-2)^{-1} + 2b(c-1)(M-1)^{-1}D + (c-1)^2] \sigma_e^4 \end{aligned}$$

Inspection of the above equation shows that the minimum MSE depends on the parameters and so minimum MSE estimation is not generally possible in practice. If we impose $b = 1$, however, we can find a conditional minimum MSE solution, which does not depend on the parameters. Solving the equation (with $b = 1$)

$$\frac{d\text{MSE}(\tilde{\sigma}_y^2)}{dc} = 2(b-1)\beta^2 \sigma_t^2 \sigma_e^2 + [4c(K-2)^{-1} + 2b(M-1)^{-1}D + 2(c-1)]\sigma_e^4 = 0$$

for c yields

$$c = c_{cmse} = (K-2)K^{-1}[1 - (M-1)^{-1}D] ; \quad K > 3$$

This minimizes $\text{MSE}(\tilde{\sigma}_y^2)$ conditionally on $b = 1$ since

$$\frac{d^2 \text{MSE}(\tilde{\sigma}_y^2)}{dc^2} = [4(K-2)^{-1} + 2]\sigma_e^4 > 0 \quad \text{for } K > 2$$

We shall call this estimator a conditional minimum MSE estimator. The estimator entails a shrinkage relative to the unbiased estimator, which is common in minimum MSE estimation (Stuart and Ord, 1991, § 17.30). The conditional estimator will be inferior to the minimum MSE solution without restriction on b , provided parameters are regarded as known.

When minimization of MSE is desired and covariate information ignored/not available, one may use (3) with (Cox and Hinkley, 1974, § 8.5) $a = a_{mse} = (K+1)^{-1}$. Note that (6) with c_{cmse} (conditional minimum MSE estimator) and $b = 1$ is not identical to (3) with $a = a_{mse}$, when $K = M$. This is in contrast to minimum variance unbiased estimation, where (6) and (3) coincide when $K = M$. It is possible that the conditional minimum MSE estimator yields a larger MSE than (3) with any value of the variance components, for example when $K = 6$ and $M = 7$ and when $K = 6$ and $M = 6$. Due to these unfavorable features of the conditional minimum MSE estimator, we mainly discuss the unbiased estimators.

Appendix D

We now derive the REML estimator for σ_y^2 . Let \mathbf{H}_M be a $(M-1) \times M$ matrix obtained from a M -dimensional Helmert-matrix (Searle, 1982, p.71) by dropping the first row, so that

$\mathbf{H}_M \mathbf{H}_M' = \mathbf{I}_{M-1}$. REML estimation can be based on the $(M+K-2)$ contrasts $\mathbf{y}^* = \mathbf{H}_K \mathbf{y} = (y_1^*, \dots, y_{K-1}^*)$ and $\mathbf{t}^* = \mathbf{H}_M \mathbf{t} = (t_1^*, \dots, t_{M-1}^*)$, where \mathbf{y} and \mathbf{t} are as given in (4) and (5). Note that (t_j^*, y_j^*) ($j = 1, \dots, K-1$) has a bivariate normal distribution with zero expectation and

variance-covariance matrix $\Sigma_t = \begin{pmatrix} \sigma_y^2 & \sigma_{yt} \\ \sigma_{ty} & \sigma_t^2 \end{pmatrix}$, while t_j^* ($j = K, \dots, M-1$) has zero mean and

variance σ_t^2 . Observations j and j' for which $j \neq j'$ are stochastically independent. Using results on the bivariate normal distribution (Stuart and Ord, 1991, §18.29) the first derivatives of the log-likelihood for \mathbf{y}^* and \mathbf{t}^* with respect to the variance-covariance parameters can be shown to be

$$\frac{d \log L}{d(\sigma_t^2)} = -\frac{1}{2\sigma_t^2(1-\rho^2)} \left\{ (K-1)(1-\rho^2) - \frac{S_1}{\sigma_t^2} + \rho \frac{CP}{\sigma_t \sigma_y} \right\} - \frac{(M-K)}{2\sigma_t^2} + \frac{S_2}{2(\sigma_t^2)^2} = 0$$

$$\frac{d \log L}{d(\sigma_y^2)} = -\frac{1}{2\sigma_t^2(1-\rho^2)} \left\{ (K-1)(1-\rho^2) - \frac{S_y}{\sigma_y^2} + \rho \frac{CP}{\sigma_t \sigma_y} \right\} = 0$$

$$\frac{d \log L}{d\rho} = \frac{1}{(1-\rho^2)} \left\{ (K-1)\rho - \frac{1}{(1-\rho^2)} \left[\rho \left(\frac{S_1}{\sigma_t^2} + \frac{S_y}{\sigma_y^2} \right) - (1+\rho^2) \frac{CP}{\sigma_t \sigma_y} \right] \right\} = 0$$

where $S_y = (y^*)' y^*$, $S_1 = (t_{(1)}^*)' t_{(1)}^*$, $S_2 = (t_{(2)}^*)' t_{(2)}^*$, and $CP = (y^*)' t_{(1)}^*$ with $t_{(1)}^* = (t_1^*, \dots, t_{K-1}^*)'$ and $t_{(2)}^* = (t_K^*, \dots, t_{M-1}^*)'$. After some slight rearrangement

$$(K-1)(1-\rho^2) = \frac{S_1}{\sigma_t^2} - \rho \frac{CP}{\sigma_t \sigma_y} - (M-K)(1-\rho^2) + \frac{S_2(1-\rho^2)}{\sigma_t^2} \quad (D1)$$

$$(K-1)(1-\rho^2) = \frac{S_y}{\sigma_y^2} - \rho \frac{CP}{\sigma_t \sigma_y} \quad (D2)$$

$$(K-1)(1-\rho^2) = \frac{S_1}{\sigma_t^2} + \frac{S_y}{\sigma_y^2} - \frac{(1+\rho^2)}{\rho} \frac{CP}{\sigma_t \sigma_y} \quad (D3)$$

Adding (D1) and (D2) and subtracting (D3) from this sum yields after some algebra

$$\begin{aligned} (K-1)(1-\rho^2) &= \frac{1-\rho^2}{\rho} \frac{CP}{\sigma_t \sigma_y} - (M-K)(1-\rho^2) + \frac{S_2(1-\rho^2)}{\sigma_t^2} \\ \hline (K-1) &= \frac{1}{\rho} \frac{CP}{\sigma_t \sigma_y} - (M-K) + \frac{S_2}{\sigma_t^2} \\ \hline (M-1) - \frac{S_2}{\sigma_t^2} &= \frac{1}{\rho} \frac{CP}{\sigma_t \sigma_y} \\ \hline \frac{(M-1)\sigma_t^2 - S_2}{\sigma_t^2} &= \frac{1}{\rho} \frac{CP}{\sigma_t \sigma_y} \\ \hline \rho &= \frac{\sigma_t^2}{(M-1)\sigma_t^2 - S_2} \frac{CP}{\sigma_t \sigma_y} \end{aligned} \quad (D4)$$

Rearranging (D1)

$$(M-1)(1-\rho^2) = \frac{S_1}{\sigma_t^2} - \rho \frac{CP}{\sigma_t \sigma_y} + \frac{S_2(1-\rho^2)}{\sigma_t^2}$$

$$(1-\rho^2) \left[\frac{(M-1)\sigma_t^2 - S_2}{\sigma_t^2} \right] = \frac{S_1}{\sigma_t^2} - \rho \frac{CP}{\sigma_t \sigma_y}$$

and using D4

$$(1-\rho^2) \left[\frac{(M-1)\sigma_t^2 - S_2}{\sigma_t^2} \right] = \frac{S_1}{\sigma_t^2} - \rho^2 \left[\frac{(M-1)\sigma_t^2 - S_2}{\sigma_t^2} \right], \text{ which yields} \quad (\text{D5})$$

$$\sigma_t^2 = (M-1)^{-1}(S_1 + S_2) \quad (\text{D6})$$

Substituting (D6) back into the denominator of the first ratio on the r.h.s. of (D4) yields

$$\rho = \frac{\sigma_t}{S_1} \frac{CP}{\sigma_y}, \text{ whence} \quad (\text{D7})$$

$$\beta = \rho \frac{\sigma_y}{\sigma_t} = \frac{CP}{S_1} \quad (\text{D8})$$

Rearranging (D8) we find

$$\rho = \beta \frac{\sigma_t}{\sigma_y} \text{ and} \quad (\text{D9})$$

$$\rho^2 = \beta^2 \frac{\sigma_t^2}{\sigma_y^2} \quad (\text{D10})$$

Inserting (D9) and (D10) in (D2)

$$(K-1) \frac{\sigma_y^2 - \beta^2 \sigma_t^2}{\sigma_y^2} = \frac{S_y}{\sigma_y^2} - \beta \frac{CP}{\sigma_y^2}$$

$$(K-1)(\sigma_y^2 - \beta^2 \sigma_t^2) = S_y - \beta^2 S_1$$

$$(K-1)\sigma_y^2 = S_y - \beta^2 S_1 + (K-1)\beta^2 \sigma_t^2$$

$$\sigma_y^2 = \beta^2 \sigma_t^2 + (K-1)^{-1}(S_y - \beta^2 S_1)$$

$$\sigma_y^2 = \beta^2 [\sigma_t^2 - (K-1)^{-1} S_1] + (K-1)^{-1} S_y \quad (\text{D11})$$

Thus, from (D6), (D8) and (D11) the REML estimator of σ_y^2 is

$$\hat{\sigma}_y^2 = \hat{\beta}^2 [\hat{\sigma}_t^2 - (K-1)^{-1} S_1] + (K-1)^{-1} S_y$$

with $\hat{\sigma}_t^2 = (M-1)^{-1} (S_1 + S_2)$ and $\hat{\beta} = \frac{CP}{S_1}$. Noting that $S_1 + S_2 = SS_M$, $S_1 = SS_1$, and

$\hat{\sigma}_e^2 = (K-2)^{-1} (S_y - \hat{\beta}^2 S_1)$ the estimator can be rewritten as

$$\hat{\sigma}_y^2 = \hat{\beta}^2 \hat{\sigma}_t^2 + (K-2)(K-1)^{-1} \hat{\sigma}_e^2$$

which is equivalent to (6) with $c = (K-2)(K-1)^{-1}$.

Appendix E

The four quantities needed for evaluating the Taylor expansion in (17) are:

- (i) $E(\hat{\beta}^2) = \beta^2 + (K-3)^{-1} \sigma_e^2 \sigma_t^{-2}$
- (ii) $\text{var}(\hat{\beta}^2) = E_t[\text{var}(\hat{\beta}^2 | \mathbf{t})] + \text{var}_t[E(\hat{\beta}^2 | \mathbf{t})]$
- (iii) $E(\hat{\sigma}_t^2) = \sigma_t^2$
- (iv) $\text{var}(\hat{\sigma}_t^2) = 2(M-1)^{-1} \sigma_t^2$

Ad (ii): With

$$\text{var}_t(\hat{\beta}^2 | \mathbf{t}) = 2SS_1^{-2} \sigma_e^4 + 4\beta^2 SS_1^{-1} \sigma_e^2,$$

$$E_t[\text{var}(\hat{\beta}^2 | \mathbf{t})] = 2(K-3)^{-1}(K-5)^{-1} \sigma_e^4 \sigma_t^{-4} + 4(K-3)^{-1} \beta^2 \sigma_e^2 \sigma_t^{-2},$$

$$E(\hat{\beta}^2 | \mathbf{t}) = \beta^2 + SS_1^{-1} \sigma_e^2, \text{ and}$$

$$\text{var}_t[E(\hat{\beta}^2 | \mathbf{t})] = \{E(SS_1^{-2}) - [E(SS_1^{-1})]^2\} \sigma_e^4 = 2(K-3)^{-2}(K-5)^{-1} \sigma_e^4 \sigma_t^{-4} \quad \text{we find}$$

$$\text{var}(\hat{\beta}^2) = 2(K-2)(K-3)^{-2}(K-5)^{-1}\sigma_e^4\sigma_t^{-4} + 4(K-3)^{-1}\beta^2\sigma_e^2\sigma_t^{-2}$$

Thus

$$T_1 = [E(\hat{\sigma}_t^2)]^2 \text{var}(\hat{\beta}^2) \sigma_y^{-4} = 2(K-2)(K-3)^{-2}(K-5)^{-1}q^2 + 4(K-3)^{-1}q(1-q) \text{ and}$$

$$T_2 = [E(\hat{\beta}^2)]^2 \text{var}(\hat{\sigma}_t^2) \sigma_y^{-4} = 2(M-1)^{-1}[(1-q) + (K-3)^{-1}q]^2 \text{ with } q = \sigma_e^2 / \sigma_y^2.$$

Interesting limiting cases:

(I) $M \rightarrow \infty \Rightarrow T_2 \rightarrow 0$ and $T_1 > 0$ is constant

(II) $\sigma_e^2 = 0$ ($q = 0$) $\Rightarrow T_1 = 0 < T_2 = 2(M-1)^{-1}$

(III) $\beta^2 \sigma_t^2 = 0$ ($q = 1$) $\Rightarrow T_1 = 2(K-2)(K-5)^{-1}(K-3)^{-2} > T_2 = 2(M-1)^{-1}(K-3)^{-2}$

(IV) Because of (II) and (III), and because T_1 and T_2 are quadratic in q , for any K and M with $M \geq K > 5$, there will always be one and only one crossing point $0 \leq q_p \leq 1$, for which $T_1 = T_2$.

The crossing point will move towards larger values of q_p when the ratio T_1/T_2 decreases for all values of q . For fixed q , the ratio T_1/T_2 will decrease as $M \rightarrow K$. Thus, for given q and K , the smallest value for T_1/T_2 will occur when $M = K$. Setting $M = K$, the ratio T_1/T_2 decreases as $K \rightarrow \infty$. Thus, the largest value for q_p occurs when $M = K$ and $K \rightarrow \infty$. We find

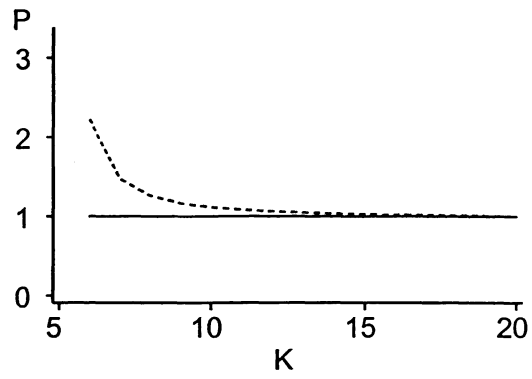
$\lim_{K \rightarrow \infty} (q_p | M = K) = 1/3$. Thus, when $q > 1/3$ we have $T_1 > T_2$ for any value of M and K

satisfying $M \geq K > 5$. For $q < 1/3$, T_1 will usually dominate T_2 , (i.e. $T_1 > T_2$) unless q is rather small.

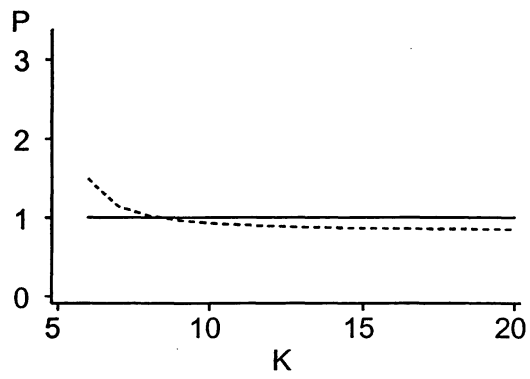
Table 1: Yield data on two wheat cultivars, one standard (t_i) and one new cultivar (y_i).

Year	75	76	77	78	79	80	81	82	83	84	85	86
t_i	4.60	13.90	5.10	16.30	4.30	9.10	17.30	6.30	19.93	23.12	17.50	16.60
y_i	-	-	-	-	-	-	16.90	6.20	24.03	22.10	16.25	8.53

$$q = 0.8$$



$$q = 0.5$$



$$q = 0.2$$

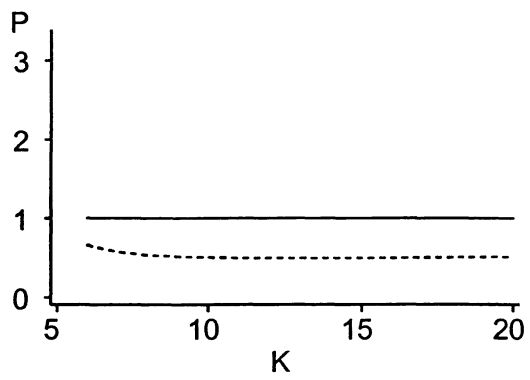


Fig.1: Plot of P vs. K for $b = 1$ and $c = c_{unb}$ and $M = 100$ and different values of q .

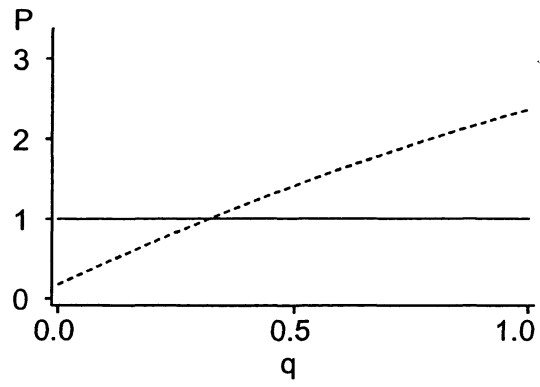
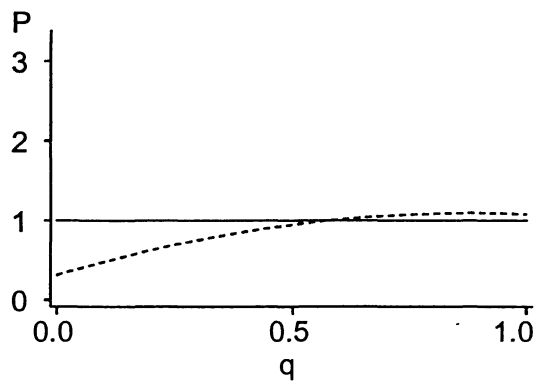
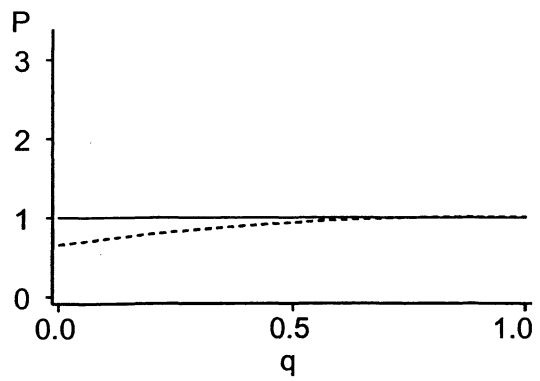
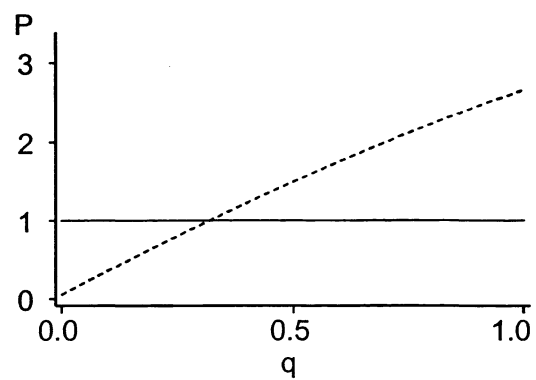
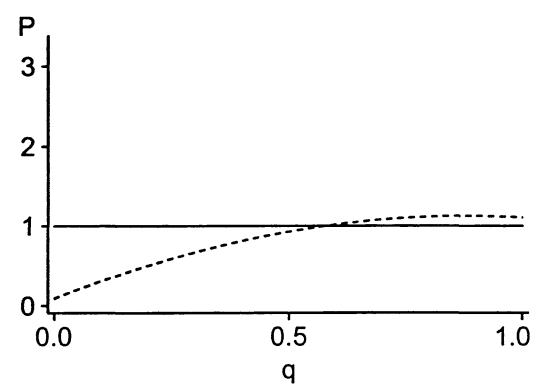
$K = 6$

 $K = 10$

 $K = 20$


Fig. 2: Plot of P vs. q for $b = 1$ and $c = c_{unb}$ and different values of K and $M = 30$.

$K = 6$



$K = 10$



$K = 20$

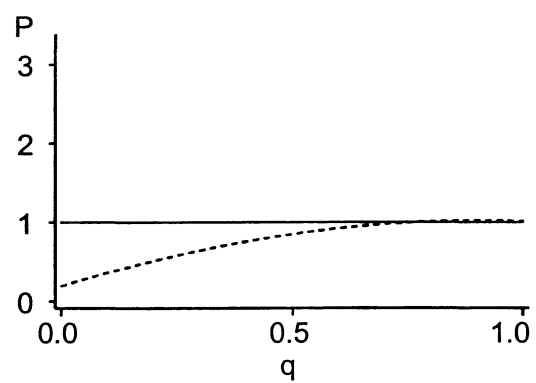
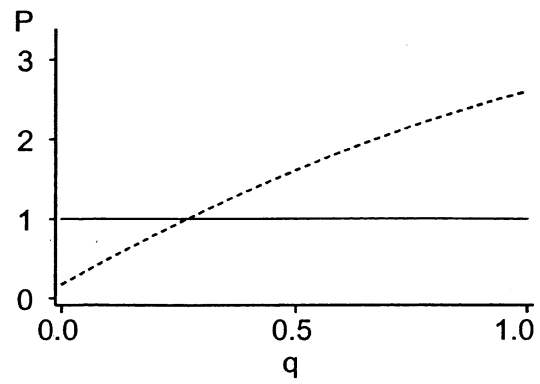
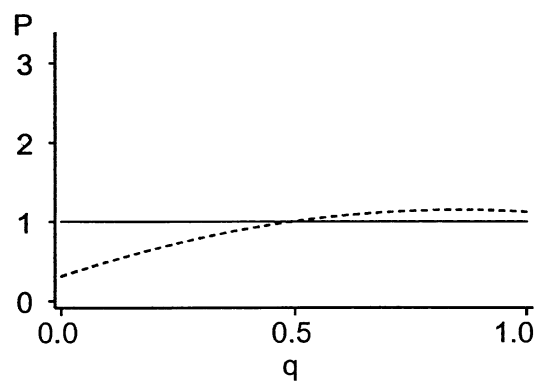


Fig. 3: Plot of P vs. q for $b = 1$ and $c = c_{unb}$ and different values of K and $M = 100$.

$K = 6$



$K = 10$



$K = 20$

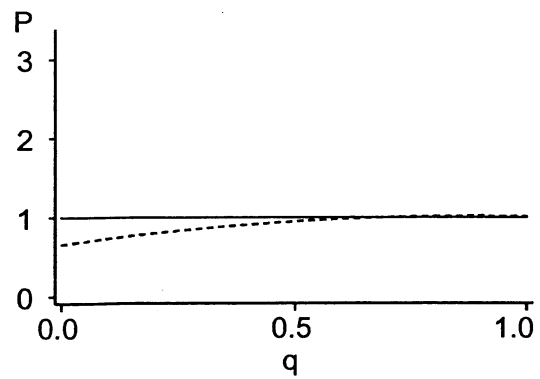


Fig. 4: Plot of P vs. q for $b = 1$ and $c = c_{reml}$ and different values of K and $M = 30$.